INTRODUCTION TO ROBUST ESTIMATION WITH APPLICATIONS

Ampai Ussawarujikulchai
Department of Mathematical Sciences
Montana State University

December 10, 1999

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

Ampai Ussawarujikulchai

This writing project has been read by the writing project director and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	Robert J. Boik Writing Project Director

Contents

1 Introduction

Estimation of unknown parameters is an important component of nearly all conventional data analyses. Statisticians are rightly concernedwith the lack of robustness that characterize many of the usual estimators. Often, the parameters to be estimated also suffer from lack of robustness. That is, a small change in the underlying distribution can result in a large change in the value of the parameter.

In this paper, robust estimation and related topics are considered, especially the influence function and its application. In section 2, robust estimation, the influence function, and the breakdown point are defined and illustrated. In section 3, estimates of standard error based on the influence function are described. In section 4, applications of the influence function are illustrated. An example of variance of trimmed means is illustrated in section 5.

Much of material in this paper was obtained from Wilcox (1997) and Staudte and Sheather (1990).

2 Robust Estimation and Properties

Location and scale measures are two types of measures that characterize a distribution. These measures are said to be robust if they are insensitive to slight changes in a distribution. The basic tools for judging robustness of the measures of location and scale are infinitesimal robustness and quantitative robustness.

2.1 Infinitesimal Robustness

An estimator or parameter is said to have infinitesimal robustness if its influence function is bounded. The influence function measures the effect on an estimator or parameter of deviation in a distribution.

Let F be a distribution function. A parameter of F, say θ, can be expressed as a functional of F:

$$\theta = T(F).$$

To define the influence function, another distribution function is needed. Let x be an arbitrary value in the support set of F and let

$$\delta_x(y) = \begin{cases} 0, & \text{if } y < x; \text{ and} \\ 1, & \text{if } y \geq x. \end{cases}$$

Note that δ_x is a degenerate distribution with probability mass one at x.

The influence function $\mathrm{IF}(x)$ of T at F is defined (Hampel, 1974) as

$$\mathrm{IF}(x) = \lim_{\epsilon \to 0} \frac{\hat{\theta} - \theta}{\epsilon}, \tag{1}$$

where

$$\hat{\theta} = T(F_{x,\epsilon}),$$

and

$$F_{x,\epsilon} = (1-\epsilon)F + \epsilon\delta_x.$$

The distribution function, $F_{x,\epsilon}$, is a mixture distribution, where an observation is randomly sampled from distribution F with probability $1-\epsilon$ and from distribution δ_x with probability ϵ.

In equation (1), IF(x) is the relative influence of x on θ when the probability of sampling from δ_x is arbitrarily close to 0. To illustrate the influence function, suppose that F has mean μ. It follows that $F_{x,\epsilon}$ has mean $(1\text{-}\epsilon)\mu + \epsilon x$. The difference between the mean of $F_{x,\epsilon}$ and the mean of F is $\epsilon(x-\mu)$. This shows that $F_{x,\epsilon}$ is similar to F when ϵ is small.

2.1.1 Influence Function of the Population Mean

Consider the influence function of the population mean, if

$$\theta = E(X), \text{ then } \hat{\theta} = (1-\epsilon)\theta + \epsilon x,$$

and

$$\frac{\hat{\theta}-\theta}{\epsilon} = x - \theta.$$

Thus, IF$(x) = x - \theta$ which does not depend on F. Also the influence function is unbounded in x. That is θ does not have infinitesimal robustness.

2.1.2 Influence Function of Quantiles

Consider the influence function of a quantile. Let x_q be the q^{th} quantile of F. That is $F(x_q) = q$. To find the influence function of a quantile, first define the q^{th} quantile as

$$x_q = \inf\{x_q; F(x_q) \geq q\}.$$

Let $f(x)$ represent the probability density function. Assume that $f(x_q) > 0$ and f is continuous at x_q.

It is shown in Appendix A that the influence function of x_q is

$$\text{IF}_q(x) = \begin{cases} \frac{q-1}{f(x_q)}, & \text{if } x < x_q; \\ 0, & \text{if } x = x_q; \text{ and} \\ \frac{q}{f(x_q)}, & \text{if } x > x_q. \end{cases} \tag{2}$$

This influence function is bounded, so the quantiles have infinitesimal robustness.

2.1.3 Influence Function of Correlation Coefficient

A principal utility of correlation is to determine the linear relationship between two random variables, X_1 and X_2. The corresponding measure of correlation is

$$\rho_{12} = \frac{\sigma_{x_1x_2}}{\sigma_{x_1}\sigma_{x_2}}.$$

The limiting value of $\rho_{12} = 0$ occurs when X_1 and X_2 are independent or uncorrelated. In contrast, the value of $\rho_{12} = 1$ occurs when X_2 is a linear function of X_1 ($X_2 = \alpha + \beta X_1$).

The conventional estimator of ρ_{12} is

$$r = \frac{\Sigma(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\Sigma(X_{1i} - \bar{X}_1)^2 \Sigma(X_{2i} - \bar{X}_2)^2}}.$$

This estimator is called the Pearson's correlation coefficient. It is a biased estimator of ρ_{12} for all ρ_{12} satisfying $|\rho_{12}| \in (0, 1)$. The bias is small when sample size is large. If $\rho_{12}= 0$, then r is unbiased and if $\rho_{12} = \pm 1$, then $r = \pm 1$.

A problem with ρ_{12} is that it lacks robustness. It is sensitive to slight changes of the marginal distribution and its estimator can be affected by outliers.

Consider the influence function of the correlation coefficient. The parameter ρ_{12} is a functional and it can be written as

$$\rho_{12} = T(F_{X_1,X_2}) = \frac{\mathrm{E}(\mathrm{X}_1\mathrm{X}_2) - \mathrm{E}(\mathrm{X}_1)\mathrm{E}(\mathrm{X}_2)}{\sqrt{[E(X_1^2) - E(X_1)^2][E(X_2^2) - E(X_2)^2]}}.$$

It is shown in Appendix B that the influence function of ρ_{12} is

$$\mathrm{IF}(\boldsymbol{\omega}) = \mathrm{z}_1\mathrm{z}_2 - \frac{1}{2}\rho_{12}(\mathrm{z}_1^2 + \mathrm{z}_2^2), \qquad (3)$$

where $\mathrm{z}_i = \frac{\omega_i - \mu_i}{\sigma_i}$. Note that $\mathrm{E}(\mathrm{IF}(\boldsymbol{\omega})) = \rho_{12} - \frac{1}{2}\rho_{12}(1 + 1) = 0$.

This influence function is not bound. That is ρ_{12} does not have infinitesimal robustness.

2.2 Quantitative Robustness

The minimum value of ϵ, for which a functional, $T(F_{x,\epsilon})$, goes to infinity as x gets large is called the breakdown point of T. To illustrate this concept, again consider

$$F_{x,\epsilon} = (1 - \epsilon)F + \epsilon\delta_x$$

which has mean $(1 - \epsilon)\mu + \epsilon x$. Thus for any $\epsilon > 0$, the mean of $F_{x,\epsilon}$ can go to infinity by increasing x.

An estimator is said to have quantitative robustness if its breakdown point is greater than 0. Also the breakdown point of any equivarient location estimator cannot exceed 0.5 (Goodall, 1983, p.357). The general idea of breakdown point is to describe the effect of a small change in F on some functional $T(F)$.

The breakdown point of the population mean is 0 because for any $\epsilon > 0$, the mean can go to infinity by allowing x to go to infinity. In contrast, the breakdown point of the median is 0.5. This is the largest possible value for a location estimator.

3 Influence Function Estimates of Standard Error

The standard error of estimate is a measure of dispersion of the distribution of the estimator, so it indicates the accuracy of the estimator.

Let $T_n = T_n(x_1, ..., x_n)$ be an estimator of a parameter θ. The standard error of an estimator T_n, is $SE[T_n] = \sqrt{Var(T_n)}$. If there exists a function $V(T, F)$ that satisfies $nVar[T_n] \to V(T, F)$, then the standard error can be approximated as $SE[T_n] = \sqrt{\frac{V(T,F)}{n}}$.

In many cases, $V(T, F) = E_F[\text{IF}(x)]^2$ therefore we can estimate $V(T, F)$ by

$$\hat{V}(T, F) = \hat{E}_{F_n}[\text{IF}(x)]^2 = \frac{1}{n}\sum_{i=1}^{n} \text{IF}^2(x_i). \tag{4}$$

Accordingly,

$$\widehat{SE}[T_n] = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} \text{IF}(x_i)^2}. \tag{5}$$

The influence function estimate of the standard error is obtained by dividing (4) by n and taking the square root. The result is given in (5). For example, the influence function of mean is $\text{IF}(x) = x - T(F)$, and an estimate of asymptotic variance is equal to $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$. Thus the influence function estimate of standard error is

$$\widehat{SE}[\bar{X}_n] = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$$

$$= \sqrt{n-1}\frac{S_n}{n},$$

where

$$S_n = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}.$$

3.1 Justification for Equation (5)

This material is taken from Staudte and Sheather (1990, p.62-63). Let $x_1, ..., x_n$ be a random sample from a population with cdf F and let F_n be corresponding empirical cdf. If a statistic $T_n = T_n(x_1, ..., x_n)$ can be written as a functional T of F_n, $T_n = T(F_n)$, where T does not depend on n, then T is a statistical functional.

Von Mises introduced statistical functionals and proposed that a form of Taylor expansion be used to approximate a given statistical functional $T(F_n)$ in order to analyze its asymptotic properties.

The influence function appears as the first derivative term in a von Mises expansion, which is an expansion for $T(F_n)$ for F_n in a neighborhood of F:

$$T(F_n) = T(F) + \int \text{IF}_{T,F}(x)d(F_n - F)(x) + R_n,$$

where R is a remainder term and satisfies $n^{1/2}R_n \to 0$. It can be show that $\int \text{IF}_{T,F}(x)dF(x) = 0$. Accordingly, the difference

$$n^{1/2}\left[T(F_n) - T(F) - \int \text{IF}_{T,F}(x)dF_n(x)\right]$$

converges to zero in probability. Hence, we have the approximation

$$n^{1/2}[T(F_n) - T(F)] \approx n^{1/2}\sum_{i=1}^{n} \text{IF}_{T,F}(X_i). \qquad (6)$$

The right hand side of equation(6) satisfies

$$E_F[\text{IF}_{T,F}(X)] = 0$$

and

$$\begin{aligned} Var_F[\text{IF}_{T,F}(X)] &= E[\text{IF}_{T,F}(X)]^2 - [E[\text{IF}_{T,F}(X)]]^2 \\ &= [E[\text{IF}_{T,F}(X)]]^2. \end{aligned}$$

Hence,

$$V(T, F) = E[\text{IF}_{T,F}(X)]^2.$$

Also, it follows from (6) that

$$T(F_n) = T(F) + \frac{1}{n}\sum_{i=1}^{n} \text{IF}_{T,F}(X) + R_n,$$

where $n^{1/2}R_n \to 0$. Furthermore, by the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \text{IF}_{T,F} \xrightarrow{\text{dist}} N[0, V(T, F)].$$

4 Applications

4.1 Standard Error of a Quantile Estimator

The influence function for the q^{th} quantile was given in (2). Using (5), the estimated standard error of the sample quantile is

$$\widehat{SE}(X_q) = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} \text{IF}_q(x_i)^2},$$

where

$$\text{IF}_q(x) = \begin{cases} \frac{q-1}{f(x_q)}, & \text{if } x < x_q; \\ 0, & \text{if } x = x_q; \text{ and} \\ \frac{q}{f(x_q)}, & \text{if } x > x_q. \end{cases}$$

Consider the q^{th} sample quantile, assuming that $[nq]$ values are smaller than x_q and $[n - nq = n(1-q)]$ values are larger than x_q. Thus, the standard error of the q^{th} sample quantile is

$$\begin{aligned}\widehat{SE}(X_q) &= \sqrt{\frac{1}{n^2}\left[\frac{nq(q-1)^2}{f(x_q)^2} + \frac{n(1-q)q^2}{f(x_q)^2}\right]},\\ &= \sqrt{\frac{1}{n^2}\left[\frac{nq(q-1)}{f(x_q)^2}\right]},\\ &= \sqrt{\frac{1}{n}\frac{q(1-q)}{f(x_q)^2}}\\ &= \frac{\sqrt{q(1-q)}}{\sqrt{n}f(x_q)}.\end{aligned}$$

The standard error of the median, for example, is $\frac{1}{2\sqrt{n}f(x_q)}$.

4.2 Standard Error of the Sample Correlation Coefficient

The influence function for the correlation coefficient was given in (3). Using (5), the estimated standard error of the sample correlation coefficient is

$$\widehat{SE}(r) = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{IF}(x_i)^2},$$

where

$$\begin{aligned}\mathrm{IF}(x_i) &= z_{1i}z_{2i} - \frac{1}{2}\rho(z_{1i}^2 + z_{2i}^2),\\ z_{1i} &= \frac{x_{1i}-\mu_1}{\sigma_1},\ z_{2i} = \frac{x_{2i}-\mu_2}{\sigma_2}.\end{aligned}$$

4.3 Variance of Vector of Trimmed Mean

The 2β - trimmed mean is defined as

$$T_{2\beta} = \mu_{2\beta} = \int_{F^{-1}(\beta)}^{F^{-1}(1-\beta)} \frac{uf(u)}{1-2\beta}du.$$

It is a symmetrically trimmed mean in which β is trimmed at each end where $\beta \in (0, \frac{1}{2})$. Let $\mathbf{X}$ be a random p - vector with cdf $F_{\mathbf{X}}$. The perturbed joint cdf of $\mathbf{X}$ is

$$F_\epsilon(\mathbf{X}) = (1-\epsilon)F_{\mathbf{X}} + \epsilon\delta_{\boldsymbol{\omega}},$$

where $\boldsymbol{\omega} \in$ support of $F_{\mathbf{x}}$ and

$$\delta_{\boldsymbol{\omega}}(\mathbf{x}) = \begin{cases} 1, & \text{if } X_i \geq \omega_i, \text{ i= 1,...,p;} \\ 0, & \text{otherwise.} \end{cases}$$

From (1), the influence function of the trimmed mean is

$$\mathrm{IF}(\boldsymbol{\omega}) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F_{\mathbf{X}})}{\epsilon},$$

where

$$T(F_{\mathbf{X}}) = \boldsymbol{\mu}_{2\beta} = \begin{pmatrix} \mu_{2\beta,1} \\ \mu_{2\beta,2} \\ \cdot \\ \cdot \\ \mu_{2\beta,p} \end{pmatrix};$$

$$T(F_\epsilon) = \begin{pmatrix} \mu_{2\beta,1,\epsilon} \\ \mu_{2\beta,2,\epsilon} \\ \cdot \\ \cdot \\ \mu_{2\beta,p,\epsilon} \end{pmatrix};$$

$$\mu_{2\beta,i} = \int_{F_i^{-1}(\beta)}^{F_i^{-1}(1-\beta)} \frac{u}{1-2\beta} dF_i(u);$$

$$\mu_{2\beta,i,\epsilon} = \int_{F_{i,\epsilon}^{-1}(\beta)}^{F_{i,\epsilon}^{-1}(1-\beta)} \frac{u}{1-2\beta} dF_{i,\epsilon}(u);$$

F_i is the marginal cdf of X_i;
$F_{i,\epsilon}$ is the perturbed marginal cdf of X_i ; and
$F_{i,\epsilon} = (1-\epsilon)F_i + \epsilon\delta_{\omega_i}$. Accordingly,

$$\mathrm{IF}(\omega_i) = \lim_{\epsilon \to 0} \frac{(1-\epsilon)}{\epsilon} \int_{F_{i,\epsilon}^{-1}(\beta)}^{F_{i,\epsilon}^{-1}(1-\beta)} \frac{u}{1-2\beta} dF_{i,\epsilon}(u) + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon}, X_{1-\beta,i,\epsilon})}(\omega_i)$$

$$-\frac{1}{\epsilon} \int_{F_i^{-1}(\beta)}^{F_i^{-1}(1-\beta)} \frac{u}{1-2\beta} dF_i(u), \qquad (7)$$

where I is an indicator function and $X_{\beta,i,\epsilon}$ is the 100β quantile of $F_{i,\epsilon}$. Note that $X_{\beta,i,\epsilon} = F_{i,\epsilon}{}^{-1}(\beta)$.

The expansion of $F_{i,\epsilon}^{-1}(\beta)$ and $F_{i,\epsilon}^{-1}(1-\beta)$ around ϵ=0 is

$$F_{i,\epsilon}^{-1}(\beta) = F_i^{-1}(\beta) + \frac{d}{d\epsilon} F_{i,\epsilon}^{-1}(\beta)\Big|_{\epsilon=0} \epsilon + O(\epsilon^2), \text{and}$$

$$F_{i,\epsilon}^{-1}(1-\beta) = F_i^{-1}(1-\beta) + \frac{d}{d\epsilon} F_{i,\epsilon}^{-1}(1-\beta)\Big|_{\epsilon=0} \epsilon + O(\epsilon^2).$$

Using Equation (19) from Appendix A, the first expansion is

$$F_{i,\epsilon}^{-1}(\beta) = x_{\beta,i} + \frac{\beta - \delta_{\omega_i}(x_{\beta,i})}{f_i(x_{\beta,i})}\epsilon + O(\epsilon^2),$$

where $x_{\beta,i} = F_i^{-1}(\beta)$.

Substituting these expansion in equation (7) yields

$$\begin{aligned} \mathrm{IF}(\omega_i) &= \lim_{\epsilon \to 0} \frac{(1-\epsilon)}{\epsilon} \int_{x_{\beta,i}+\frac{\beta-\delta_{\omega_i}(x_{\beta,i})}{f_i(x_{\beta,i})}\epsilon}^{x_{1-\beta,i}+\frac{\beta-\delta_{\omega_i}(x_{1-\beta,i})}{f_i(x_{1-\beta,i})}\epsilon} \frac{u}{1-2\beta} dF_i(u) + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon}, X_{1-\beta,i,\epsilon})}(\omega_i) \\ & -\frac{1}{\epsilon}\int_{x_{\beta,i}}^{x_{1-\beta,i}} \frac{u}{1-2\beta} dF_i(u). \end{aligned}$$

Let $a = x_{\beta,i}$, $b = \frac{\beta-\delta_{\omega_i}(x_{\beta,i})}{f_i(x_{\beta,i})}$, $c = x_{1-\beta,i}$ and $d = \frac{(1-\beta)-\delta_{\omega_i}(x_{1-\beta,i})}{f_i(x_{1-\beta,i})}$. Then,

$$\begin{aligned} \mathrm{IF}(\omega_i) &= \lim_{\epsilon \to 0} \frac{(1-\epsilon)}{\epsilon} \int_{a+b\epsilon}^{c+d\epsilon} \frac{uf(u)}{1-2\beta} du + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon}, X_{1-\beta,i,\epsilon})}(\omega_i) - \frac{1}{\epsilon}\int_a^c \frac{uf(u)}{1-2\beta} du, \\ &= \lim_{\epsilon \to 0} \frac{(1-\epsilon)}{\epsilon} \int_c^{c+d\epsilon} \frac{uf(u)}{1-2\beta} du - \frac{1}{\epsilon}\int_a^{a+b\epsilon} \frac{uf(u)}{1-2\beta} du + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon}, X_{1-\beta,i,\epsilon})}(\omega_i) \\ & -\int_{a+b\epsilon}^{c+d\epsilon} \frac{uf(u)}{1-2\beta} du, \\ &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_c^{c+d\epsilon} \frac{uf(u)}{1-2\beta} du - \frac{1}{\epsilon}\int_a^{a+b\epsilon} \frac{uf(u)}{1-2\beta} du + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon}, X_{1-\beta,i,\epsilon})}(\omega_i) - \mu_{2\beta,i}. \end{aligned}$$

Let $g(u) = uf(u)$ and by L'Hospital's rule,

$$\begin{aligned} \frac{d}{d\epsilon}\int_c^{c+d\epsilon} \frac{g(u)}{1-2\beta} du &= \frac{d}{d\epsilon} \frac{G[c+d\epsilon]-G(c)}{1-2\beta}, \\ &= \frac{g[c+d\epsilon]d}{1-2\beta}. \end{aligned}$$

Therefore,

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_c^{c+d\epsilon} \frac{uf(u)}{1-2\beta} du - \frac{1}{\epsilon}\int_a^{a+b\epsilon} \frac{uf(u)}{1-2\beta} du = \frac{cf(c)d - af(a)b}{1-2\beta}.$$

Now substitute a, b, c, and d back in to the equation for the influence function to obtain

$$\begin{aligned} \mathrm{IF}(\omega_i) &= \frac{x_{1-\beta,i}[(1-\beta)+\delta_{\omega_i}(x_{1-\beta,i})] - x_{\beta,i}[\beta+\delta_{\omega_i}(x_{\beta,i})]}{1-2\beta} \\ & + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon}, X_{1-\beta,i,\epsilon})}(\omega_i) - \mu_{2\beta,i}. \end{aligned} \tag{8}$$

Consider the Winsorization of a random variable. It consists of setting

$$W_{2\beta}(x) = \begin{cases} x_\beta, & \text{if } x < x_\beta; \\ x, & \text{if } x_\beta \le x \le x_{1-\beta}; \text{ and} \\ x_{1-\beta}, & \text{if } x > x_{1-\beta}. \end{cases} \qquad (9)$$

The Winsorized mean is

$$\begin{aligned} E(W_{2\beta}) = \mu_{w2\beta,i} &= \int_{-\infty}^{\infty} W_{2\beta}(x) f_i(x)dx, \\ &= \beta x_\beta + (1-2\beta)\mu_{2\beta} + \beta x_{1-\beta}. \end{aligned}$$

Accordingly,

$$\mu_{2\beta} = \frac{\mu_{w2\beta} - \beta x_\beta - \beta x_{1-\beta}}{1-2\beta}. \qquad (10)$$

Now substitute $\mu_{2\beta}$ from equation (10) into equation (8). The result is

$$\begin{aligned} \text{IF}(\omega_i) &= \frac{x_{1-\beta,i}[(1-\beta)+\delta_{\omega_i}(x_{1-\beta,i})] - x_{\beta,i}[\beta + \delta_{\omega_i}(x_{\beta_i})]}{1-2\beta} + \frac{\omega_i}{1-2\beta} I_{(X_{\beta,i,\epsilon},X_{1-\beta,i,\epsilon})}(\omega_i) \\ &\quad - \frac{\mu_{\omega_{2\beta,i}}}{1-2\beta} + \frac{\beta x_{\beta,i}}{1-2\beta} + \frac{\beta x_{1-\beta,i}}{1-2\beta}, \\ &= \frac{x_{1-\beta,i} - x_{1-\beta,i}\delta_{\omega_i}(x_{1-\beta,i}) + x_{\beta,i}\delta_{\omega_i}(x_{\beta,i}) + \omega_i I_{(X_{\beta,i,\epsilon},X_{1-\beta,i,\epsilon})}(\omega_i) - \mu_{\omega_{2\beta,i}}}{1-2\beta}, \\ &= \frac{x_{1-\beta,i}(1-\delta\omega_i(x_{1-\beta,i})) + x_{\beta,i}\delta\omega_i(x_{\beta,i}) + \omega_i I_{(X_{\beta,i,\epsilon},X_{1-\beta,i,\epsilon})}(\omega_i) - \mu_{\omega_{2\beta,i}}}{1-2\beta}. \end{aligned}$$

Therefore, the influence function for the trimmed mean is

$$\text{IF}(\omega_i) = \begin{cases} \frac{x_{\beta,i}-\mu_{w_2\beta,i}}{1-2\beta}, & \text{if } \omega_i < x_{\beta,i} \ ; \\ \frac{\omega_i - \mu_{w_{2\beta,i}}}{1-2\beta}, & \text{if } \omega_i \in (x_{\beta,i}, x_{1-\beta,i}) \ ; \text{ and} \\ \frac{x_{1-\beta,i}-\mu_{w2\beta,i}}{1-2\beta}, & \text{if } \omega_i > x_{1-\beta,i}. \end{cases} \qquad (11)$$

Using (5) to estimate the covariance matrix of sample trimmed means result is

$$\begin{aligned} \widehat{Var}(\widehat{\mu}_{2\beta}) &= \frac{1}{n^2}\sum_{i=1}^{n} \text{IF}(\mathbf{X}_i)[\text{IF}(\mathbf{X}_i)]', \\ &= \frac{1}{n^2}\sum_{i=1}^{n} (\mathbf{W}_{2\beta,i} - \overline{\mathbf{W}}_{2\beta,i})(\mathbf{W}_{2\beta,i} - \overline{\mathbf{W}}_{2\beta,i})', \end{aligned} \qquad (12)$$

where $\overline{\mathbf{W}}_{2\beta,i}$ is the Winsorized sample mean,

$$\overline{\mathbf{W}}_{2\beta,i} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{W}_{2\beta,i}, \qquad (13)$$

and $\mathbf{W}_{2\beta,i}$ is the vector $\begin{pmatrix} W_{2\beta,1}(x_{i1}) \\ W_{2\beta,2}(x_{i2}) \\ \cdot \\ \cdot \\ W_{2\beta,p}(x_{ip}) \end{pmatrix}$. In practice, sample quantile are substituted for population quantiles to compute $\mathbf{W}_{2\beta,i}$. When there is no trimming, then equation (12) simplifies to

$$\widehat{Var}(\widehat{\mu}_{2\beta}) = \frac{(n-1)\mathbf{S}}{n^2},$$

where $\mathbf{S} = \frac{\mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{X}}{n-1}$, $\mathbf{X}$ is the $n \times p$ matrix of responses and $\mathbf{H} = \mathbf{1}_n \frac{1}{n} \mathbf{1}_n{}'$ but $\frac{1}{n}\mathbf{S}$ is used because it is unbiased. Accordingly, to be consistent with how the standard error of the sample mean is usually estimated,

$$\widehat{Var}(\widehat{\mu}_{2\beta,i}) = \frac{1}{n(n-1)(1-2\beta)^2} \sum_{i=1}^{n} (\mathbf{W}_{2\beta,i} - \overline{\mathbf{W}}_{2\beta,i})(\mathbf{W}_{2\beta,i} - \overline{\mathbf{W}}_{2\beta,i})'$$

will be used rather than equation (12). The quantity

$$\mathbf{S}_{w2\beta} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{W}_{2\beta,i} - \overline{\mathbf{W}}_{2\beta,i})(\mathbf{W}_{2\beta,i} - \overline{\mathbf{W}}_{2\beta,i})', \tag{14}$$

is called the sample Winsorized variance. Accordingly,

$$\widehat{Var}(\widehat{\mu}_{2\beta,i}) = \frac{1}{n(1-2\beta)^2} \mathbf{S}_{w2\beta}. \tag{15}$$

5 Trimmed Mean Example

Bernard G. Greenberg (1953) reported data which consisted of the ages in months and the corresponding heights in centimeters of children from a private school.

In order to compute the covariance of sample trimmed means, first order data in Table 5.1 from the smallest to the largest separately. Using equation (9) compute the Winsorized random variable which $\beta = 0.2$. Compute the Winsorized sample means by equation (13); $\overline{\mathbf{W}}_{2\beta,x_1} = 126.83$ for age and $\overline{\mathbf{W}}_{2\beta,x_2} = 143.87$ for height. Using equation (14), $\mathbf{S}_{w2\beta}$ is compute to be

$$\mathbf{S}_{w2\beta} = \begin{pmatrix} 70.8530 & 31.8186 \\ 31.8186 & 39.9504 \end{pmatrix}$$

and using the equation (15), covariance matrix of the sample trimmed mean is

$$\widehat{Var}\begin{pmatrix} \widehat{\mu}_{2\beta,1} \\ \widehat{\mu}_{2\beta,2} \end{pmatrix} = \begin{pmatrix} 10.9342 & 4.9103 \\ 4.9103 & 6.1652 \end{pmatrix}.$$

Furthermore, the Winsorized correlation coefficient is $r_w = \frac{4.9103}{\sqrt{10.9342(6.1652)}} = 0.59805$.

In order to compare these robust statistics with the usual statistics, the arithmetic means were compute, $\overline{x}_1 = 126.83$ for age and $\overline{x}_2 = 144.54$ for height.

Table 1: Table 5.1 Age and Height of Children in a Private School.

Child	Age (months)	Height (Cm)
1	109	137.6
2	113	147.8
3	115	136.8
4	116	140.7
5	119	132.7
6	120	145.4
7	121	135.0
8	124	133.0
9	126	148.5
10	129	148.3
11	130	147.5
12	133	148.8
13	134	133.2
14	135	148.7
15	137	152.0
16	139	150.6
17	141	165.3
18	142	149.9

These results show that the arithmetic mean and the Winsorized mean of age are the same. The arithmetic mean and the Winsorized mean of height are quite similar. The usual sample correlation coefficient, $r = 0.6077$. These results show that the usual statistics and the robust statistics of this example are very similar. Consider the plot of height (cm.) versus age (month) in Figure 5.1. The child at age 141 months seems to be an outlier of this data set, however it is not an extreme departure from the other points. So it is not a serious problem. The usual statistics are not affected much by this point. Therefore both the usual statistics and the robust statistics are almost the same in this example. For comparison, a scatter plot of the Winsorized data appear in Figure 5.2. As noted, one of the (x, y) coordinates represents form data points.

Figure 5.1: Plot of Height (cm.) versus Age (months) of Children in a Private School

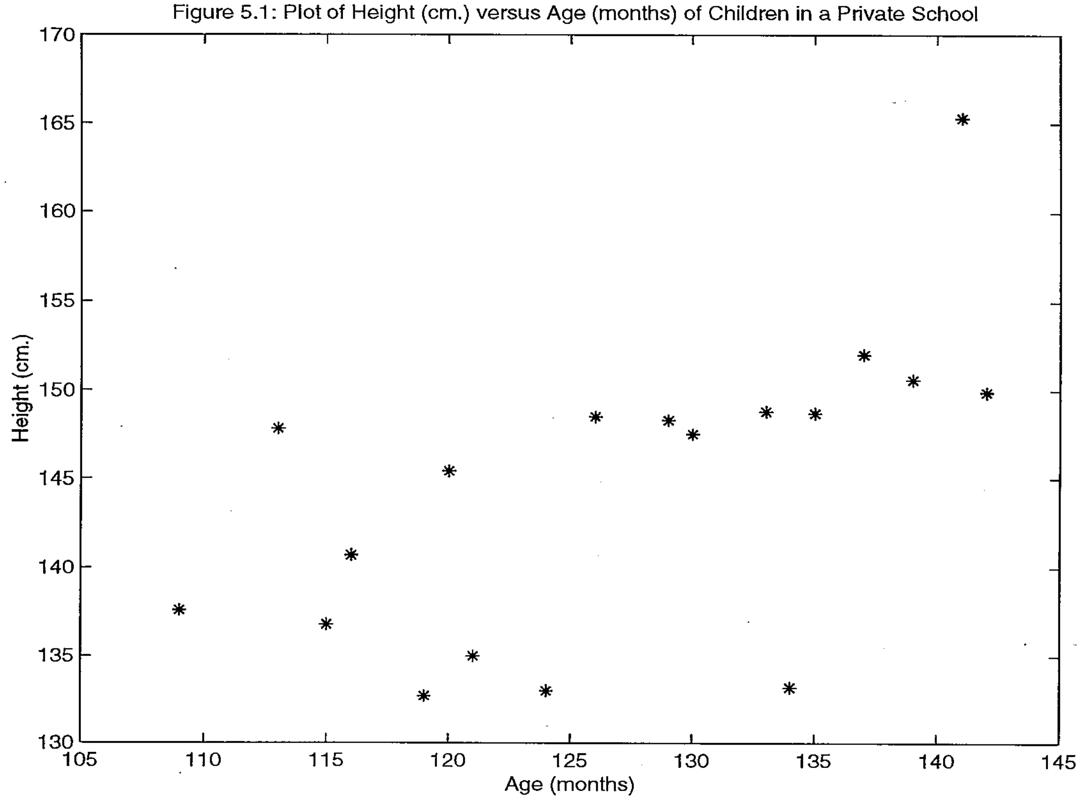
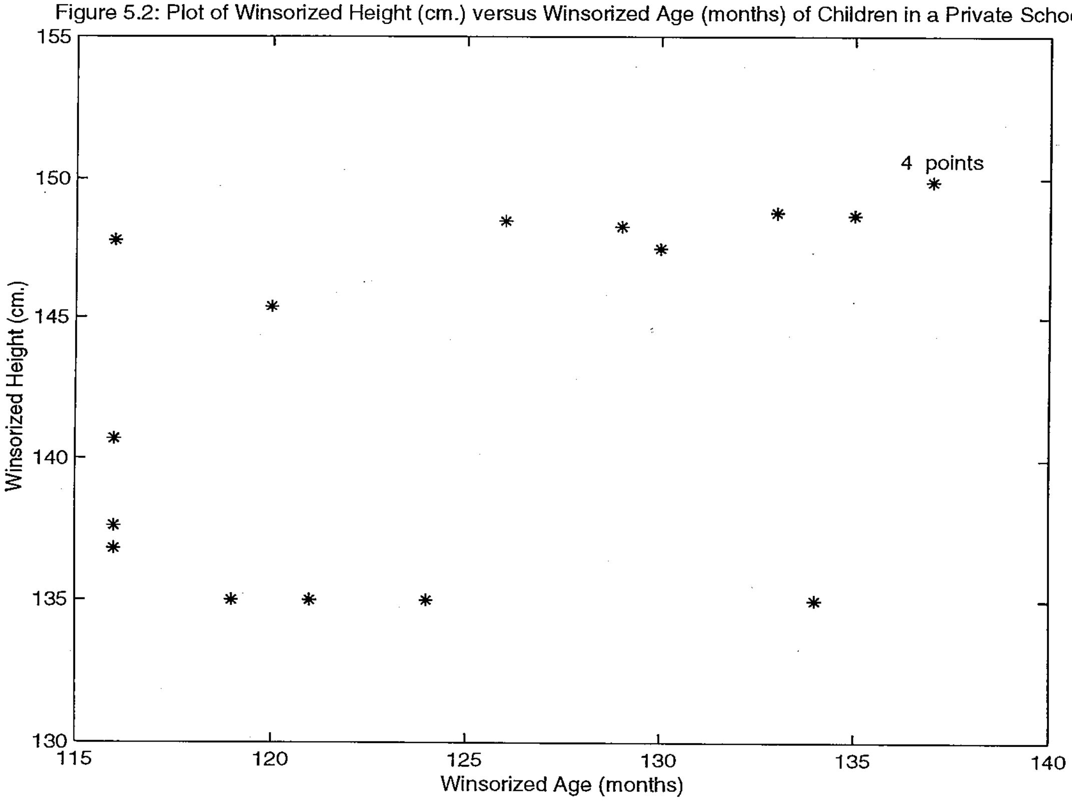


Figure 5.2: Plot of Winsorized Height (cm.) versus Winsorized Age (months) of Children in a Private School



6 APPENDIX

6.1 APPENDIX A

The Influence Function for the q^{th} quantile.

To derive the influence function for the q^{th} quantile, denote the q^{th} quantile of F by x_q. That is,

$$F(x_q) = q \text{ and } F_x^{-1}(q) = x_q.$$

The influence function $\text{IF}(x)$ of T at F is defined as

$$\text{IF}(x) = \lim_{\epsilon \to 0} \frac{T(F_{x,\epsilon}) - T(F)}{\epsilon}.$$

Therefore the influence function for the q^{th} quantile is

$$\text{IF}(x) = \lim_{\epsilon \to 0} \frac{F_{x,\epsilon}^{-1}(q) - F_x^{-1}(q)}{\epsilon}. \tag{16}$$

In case I, it is assumed that $x \neq x_q$. Begin by expanding $F_{x,\epsilon}^{-1}(q)$ around $\epsilon = 0$. The expansion is

$$F_{x,\epsilon}^{-1}(q) = F_x^{-1}(q) + \frac{d}{d\epsilon} F_{x,\epsilon}^{-1}(q)\Big|_{\epsilon=0} \epsilon + O(\epsilon^2).$$

Now,substitute this expansion in equation (16). That is

$$\text{IF}(x) = \lim_{\epsilon \to 0} \frac{F_x^{-1}(q) + \frac{d}{d\epsilon} F_{x,\epsilon}^{-1}(q)\Big|_{\epsilon=0} \epsilon + O(\epsilon^2) - F_x^{-1}(q)}{\epsilon}. \tag{17}$$

To find $\frac{d}{d\epsilon} F_{x,\epsilon}^{-1}(q)$, use

$$\begin{aligned} q &= F_{x,\epsilon}[F_{x,\epsilon}^{-1}(q)] \qquad (18) \\ &= (1-\epsilon) F_x[F_{x,\epsilon}^{-1}(q)] + \epsilon \delta_x[F_{x,\epsilon}^{-1}(q)]. \end{aligned}$$

Take the derivative of both sides of equation (17) with respect to ϵ and evaluate the derivative at $\epsilon = 0$. The derivative of the left-hand side is 0. Accordingly,

$$\frac{dq}{d\epsilon} = 0 = -F_x[F_x^{-1}(q)] + (1-\epsilon) f_x[F_{x,\epsilon}^{-1}(q)] \frac{d}{d\epsilon} F_{x,\epsilon}^{-1}(q) + \delta_x[F_{x,\epsilon}^{-1}(q)],$$

using $\frac{d}{d\epsilon}\delta_x(y) = 0$ for $x \neq y$. Thus,

$$\begin{aligned} \frac{d}{d\epsilon} F_{x,\epsilon}^{-1}(q)\Big|_{\epsilon=0} &= \frac{F_x[F_{x,\epsilon}^{-1}(q)] - \delta_x[F_{x,\epsilon}^{-1}(q)]}{(1-\epsilon) f_x[F_{x,\epsilon}^{-1}(q)]}\Big|_{\epsilon=0} \qquad (19) \\ &= \frac{F_x[F_x^{-1}(q)] - \delta_x[F_x^{-1}(q)]}{f_x[F_x^{-1}(q)]} \\ &= \frac{q - \delta_x(x_q)}{f_x(x_q)}. \end{aligned}$$

Now replace $\frac{d}{d\epsilon}F_{x,\epsilon}^{-1}(q)\Big|_{\epsilon=0}$ of equation (19) in the definition of IF(x) (equation 17) to obtain

$$\begin{aligned}
\text{IF}(x) &= \lim_{\epsilon\to 0}\frac{F_x^{-1}(q)+\epsilon(\frac{q-\delta_x(x_q)}{f_x(x_q)}+O(\epsilon^2)-F_x^{-1}(q))}{\epsilon}\\
&= \frac{q-\delta_x(x_q)}{f_x(x_q)}\\
&= \begin{cases}\frac{q-1}{f(x_q)}, & \text{if } x<x_q\text{; and}\\ \frac{q}{f(x_q)}, & \text{if } x>x_q.\end{cases}
\end{aligned}$$

In case II, it is assumed that $x = x_q$. It can be shown that

$$F_{x,\epsilon}^{-1}(q) = x_q,$$

for any $\epsilon > 0$. By the definition of IF(x), it follows that

$$\frac{F_x^{-1}(q)-F_x^{-1}(q)}{\epsilon} = \frac{0}{\epsilon} = 0, \text{ for all } \epsilon > 0,$$

therefore IF(x_q)=0.

6.2 APPENDIX B

Influence function of Correlation Coefficient

Let $\mathbf{X}$ be a $p \times 1$ random vector with cdf $F_{\mathbf{X}}$. The perturbed joint cdf of $\mathbf{X}$ is

$$F_\epsilon = (1-\epsilon)F_{\mathbf{X}} + \epsilon\delta_{\boldsymbol{\omega}},$$

where $\boldsymbol{\omega} \in$ support of $F_{\mathbf{X}}$ and

$$\delta_{\boldsymbol{\omega}}(\mathbf{x}) = \begin{cases} 1, & \text{if } X_i \geq \omega_i, \text{ i= 1,...,p;} \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$T(F_{\mathbf{X}}) = \rho_{12} = \frac{\mathrm{E(X_1X_2) - E(X_1)E(X_2)}}{\sqrt{[E(X_1^2) - E(X_1)^2][E(X_2^2) - E(X_2)^2]}},$$

for p=2.

From (1), the influence function of the correlation coefficient is

$$\begin{aligned}\mathrm{IF}(\boldsymbol{\omega}) &= \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F_x)}{\epsilon} \\ &= \frac{\frac{\mathrm{E_\epsilon(X_1X_2) - E_\epsilon(X_1)E_\epsilon(X_2)}}{\sqrt{[E_\epsilon(X_1^2) - E_\epsilon(X_1)^2][E_\epsilon(X_2^2) - E_\epsilon(X_2)^2]}} - \frac{\mathrm{E(X_1X_2) - E(X_1)E(X_2)}}{\sqrt{[E(X_1^2) - E(X_1)^2][E(X_2^2) - E(X_2)^2]}}}{\epsilon},\end{aligned}$$

where

$$\begin{aligned}\mathrm{E_\epsilon(X_1X_2)} &= \int\int X_1X_2 dF_\epsilon \\ &= (1-\epsilon)\mathrm{E(X_1X_2)} + \epsilon\omega_1\omega_2 \\ &= (1-\epsilon)(\sigma_{12} + \mu_1\mu_2) + \epsilon\omega_1\omega_2; \\ \mathrm{E_\epsilon(X_1)} &= (1-\epsilon)\mathrm{E(X_1)} + \epsilon\omega_1; \\ \mathrm{E_\epsilon(X_2)} &= (1-\epsilon)\mathrm{E(X_2)} + \epsilon\omega_2; \\ \mathrm{E_\epsilon(X_1^2)} &= (1-\epsilon)\mathrm{E(X_1^2)} + \epsilon\omega_1^2 \\ &= (1-\epsilon)(\sigma_1^2 + \mu_1^2) + \epsilon\omega_1^2; \text{ and} \\ \mathrm{E_\epsilon(X_2^2)} &= (1-\epsilon)\mathrm{E(X_2^2)} + \epsilon\omega_2^2 \\ &= (1-\epsilon)(\sigma_2^2 + \mu_2^2) + \epsilon\omega_2^2.\end{aligned}$$

The quantity $T(F_\epsilon) - T(F)$ simplifies to

$$T(F_\epsilon) - T(F) = \frac{\sigma_{12} - \epsilon\sigma_{12} + (\epsilon - \epsilon^2)(\omega_1 - \mu_1)(\omega_2 - \mu_2)}{\sqrt{[\sigma_1^2 - \epsilon\sigma_1^2 + (\epsilon - \epsilon^2)(\omega_1 - \mu_1)^2][\sigma_2^2 - \epsilon\sigma_2^2 + (\epsilon - \epsilon^2)(\omega_2 - \mu_2)^2]}} - \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}},$$

$$= \frac{\rho_{12} - \epsilon\rho_{12} + (\epsilon - \epsilon^2)z_1 z_2}{\sqrt{[1 - \epsilon + (\epsilon - \epsilon^2)z_1^2][1 - \epsilon + (\epsilon - \epsilon^2)z_2^2]}} - \rho_{12}.$$

where $z_1 = \frac{\omega_1 - \mu_1}{\sigma_1}$, and $z_2 = \frac{\omega_2 - \mu_2}{\sigma_2}$. Note that

$$\begin{aligned} \mathrm{IF}(\omega) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} &= \frac{d}{d\epsilon}(T(F) - \rho_{12})\Big|_{\epsilon=0} \\ &= \frac{-\rho_{12} + (1 - 2\epsilon)z_1 z_2}{1} - \rho_{12}[\frac{-1 + z_1^2}{2} + \frac{-1 + z_2^2}{2}] \\ &= z_1 z_2 - \frac{1}{2}\rho_{12}(z_1^2 + z_2^2). \end{aligned}$$

Note that $E(\mathrm{IF}(\omega)) = \rho_{12} - \frac{1}{2}\rho_{12}(1 + 1) = 0$.

7 References

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics.* New York: Wiley.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis.* New York: Wiley.

Huber, P. J. (1981). *Robust Statistics.* New York: Wiley.

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

Staudte, R. G. & Sheather, S. J. (1990). *Robust Estimation and Testing.* New York: Wiley.

Wilcox, R. R. (1997). *Introduction to Robust Estimation and Hypothesis Testing.* San Diago CA: Academic Press.

Wilcox, R. R. (1998). The Goal and Strategies of Robust Methods. *British Journal of Mathematical and Statistical Psychology,* **51**, 1–39.